

VU Research Portal

Exact algorithms for pairwise protein structure alignment

Wohlers, I.

2012

document version

Publisher's PDF, also known as Version of record

[Link to publication in VU Research Portal](#)

citation for published version (APA)

Wohlers, I. (2012). *Exact algorithms for pairwise protein structure alignment*. [PhD-Thesis – Research external, graduation internal, Vrije Universiteit Amsterdam]. CWI.

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal ?

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

E-mail address:

vuresearchportal.ub@vu.nl

Samenvatting

Eiwitten zijn de moleculaire machines van de cel en hebben hier vele verschillende functies. De specifieke functie van een eiwit komt voort uit zijn driedimensionale structuur; dit verband staat bekend als het structuur-functie paradigma.

Een eiwit is een gevouwen keten van aminozuurresiduen die verbonden zijn door peptidbindingen. De eiwitstructuur heeft een reguliere hoofdketen en verschillende aminozuurszijgroepen. De structuur van een eiwit wordt beschreven op vier niveaus: primair, secundair, tertiair en quaternair. De primaire structuur beschrijft de volgorde van de aminozuurresiduen. De secundaire structuur omvat eiwitmotieven zoals α -helices en β -sheets. De tertiaire structuur wordt bepaald door precieze positie van atomen in de structuur. De quaternaire structuur beschrijft de compositie van meerdere aminozuurketens.

Structurele overeenkomsten tussen eiwitten komen zeer vaak voor. Dit is om verschillende redenen. Ten eerste is het aantal mogelijke structurele configuraties, ook wel eiwitvouwingen genoemd, beperkt. Vervolgens komen zogenaamde domeinen, dit zijn delen van eiwitten die evolutionaire entiteiten zijn, voor in vele verschillende eiwitten omdat ze zijn verwijderd, toegevoegd, verwisseld, gemuteerd, *etc.* Daarnaast is er een hoge evolutionaire druk op het behouden van eiwitfunctie, en dus op het behouden van eiwitstructuur. Homologe eiwitten die dus een gezamenlijke voorouder delen, hebben daarom veelal een vergelijkbare structuur. Ten slotte kan convergente evolutie leiden tot soortgelijke structuren voor soortgelijke functies.

Door het verband tussen structuur en functie kunnen structurele overeenkomsten ons helpen om meer te weten te komen over evolutionaire en biologische taken van eiwitten. Om dit te doen moeten we de structurele overeenkomsten op een betrouwbare manier kunnen detecteren. Hiervoor gebruiken we experimentele data die de driedimensionale locatie van elk atoom in de eiwitstructuur beschrijft. Twee eiwitten worden beschouwd als structureel vergelijkbaar als de configuratie van de eiwithoofdketen soortgelijk is. Dit houdt in dat we enkel één atoom per aminozuurresidu gebruiken. De twee resulterende ketens met de representatieve atomen worden dan met elkaar vergeleken. Middels deze vergelijking verkrijgen we een één-op-één mapping tussen structureel equivalente residuen in de twee eiwitten. Deze mapping staat bekend als structure alignment, welke gebruikt kan worden om de gelijkenis van twee eiwitten door het toewijzen van een gelijkenisscore te kwantificeren. Gegeven de alignment van de eiwitten kunnen we de twee structuren superponeren in de driedimensionale ruimte om een visuele indruk van hun gelijkenis te verkrijgen.

Met behulp van optimalisatietechnieken beogen wij de beste structure alignment te vinden. Hiervoor zijn twee stappen van belang. Ten eerste moeten we een scoringsschema definiëren waarmee biologisch correcte alignment maximaal scoren. Ten tweede gebruiken we een algoritme dat de alignment vindt met de optimale score. Voor biologisch relevante scoringsschema's is het vinden van de optimale structure alignment moeilijk; het is vaak vermoedelijk een *NP*-moeilijk probleem. Als gevolg hiervan zijn veruit de meeste structure alignment algoritmes heuristisch van aard. Daarbij komt dat iedere heuristiek gebruik maakt van zijn eigen scoringsschema. Er is op dit moment geen overeenstemming over welk algoritme of scoringsschema het beste is.

Onze bijdrage aan het structure alignment probleem is tweeledig. Ten eerste beschrijven wij het probleem met behulp van wiskundige modellen en ontwerpen exacte algoritmen om deze op te lossen. Hiervoor formuleren we generieke geheeltallige lineaire programma's voor structure alignment gebaseerd op inter-residu afstandsmatrices. Onze modellen plaatsen een groot aantal bestaande structure alignment methodes in een gemeenschappelijk kader. Ten slotte vinden we de optimale oplossing van deze modellen door exacte algoritmes hiervoor te ontwerpen. Een inter-residu afstandsmatrix is een sequentiële toewijzing van een deelverzameling van afstandsmatrix rijen en kolommen van een matrix naar een deelverzameling van afstandsmatrix rijen en kolommen van de andere matrix. Deze toewijzing moet de totale score voor gepaarde inter-residu afstanden maximaliseren. Een exact algoritme voor dit probleem zal een mapping van de maximum scores teruggeven, of, als deze niet wordt gevonden binnen de gestelde tijdslimiet, de grenzen van de maximum score. We gebruiken technieken uit combinatorische optimalisatie voor onze exacte algoritmes, zoals: integer linear programming, Lagrangian relaxation, branch-and-bound en branch-and-cut.

Onze tweede bijdrage is de toepassing van onze algoritmes op problemen die enkel kunnen worden opgelost met behulp van exacte algoritmes. Bijvoorbeeld berekenen wij bewijsbaar betere alignments, verkrijgen wij kwaliteitsgaranties op alignments en kwantificeren wij eiwitgelijkenissen. Verder evalueren wij heuristische algoritmen en vergelijken wij op een nauwgezette manier verschillende scoringsschemas. Ten slotte stellen wij deze diensten beschikbaar aan structuurbiologen via een webserver.

Kort samengevat draagt het werk in dit proefschrift getiteld *Exacte algoritmen voor paarsgewijze eiwit structure alignment* bij aan het verbeteren van algoritmes en scoringsschemas voor eiwit structure alignment.